JOURNAL OF VIROLOGY, Oct. 2001, p. 8917–8926 Vol. 75, No. 19
0022-538X/01/$04.00+0 DOI: 10.1128/JVI.75.19.8917–8926.2001

HERV-K(OLD): Ancestor Sequences of the Human Endogenous Retrovirus Family HERV-K(HML-2)

KATRIN REUS,[1] JENS MAYER,[1,2] MARLIES SAUTER,[3] HANS ZISCHLER,[4] NIKOLAUS MÜLLER-LANTZSCH,[3] AND ECKART MEESE[1]*

Institut für Humangenetik[1] and Institut für Medizinische Mikrobiologie und Hygiene, Abteilung Virologie,[3] Universitätskliniken des Saarlandes, Homburg/Saar, and Primate Genetics, German Primate Center, Göttingen,[4] Germany, and Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania[2]

Received 26 February 2001/Accepted 19 June 2001

Sequences homologous to the human endogenous retrovirus (HERV) family HERV-K(HML-2) are present in all Old World primate species. A previous study showed that a central region of the HERV-K(HML-2) *gag* genes in *Hominoidea* species displays a 96-bp deletion compared to the *gag* genes in lower Old World primates. The more ancient HERV-K(HML-2) sequences present in lower Old World primates were apparently not conserved during hominoid evolution, as opposed to the deletion variants. To further clarify the evolutionary origin of the HERV-K(HML-2) family, we screened GenBank with the 96-bp *gag*-sequence characteristic of lower Old World primates and identified, to date, 10 human sequence entries harboring either full-length or partially deleted proviral structures, probably representing remnants of a more ancient HERV-K(HML-2) variant. The high degree of mutations demonstrates the long-time presence of these HERV-K(OLD) proviruses in the genome. Nevertheless, they still belong to the HML-2 family as deduced from dot matrix and phylogenetic analyses. We estimate, based on the family ages of integrated *Alu* elements and on long terminal repeat (LTR) divergence data, that the average age of HERV-K(OLD) proviruses is ca. 28 million years, supporting an integration time before the evolutionary split of *Hominoidea* from lower Old World primates. Analysis of HERV-K(OLD) LTR sequences led to the distinction of two subgroups, both of which cluster with LTRs belonging to an evolutionarily older cluster. Taken together, our data give further insight into the evolutionary history of the HERV-K(HML-2) family during primate evolution.

Integrations of different exogenous retroviral sequences into the germ line occurred frequently during evolution and gave rise to several families of endogenous retroviruses in the genomes of some invertebrate and all vertebrate families. After provirus insertion, retrotranspositional events in a retrovirus-like fashion may have increased the copy numbers of particular families. The recent analysis of the draft sequence shows that ca. 8% of the human genome is composed of retrovirus-like elements (8). Several distinct human endogenous retrovirus (HERV) families with copy numbers from 1 to 1,000 can be defined (36). Mutations and deletions rendered many of these HERVs unable to produce functional proteins, and thus they are replication defective, although many remained transcriptionally active. Unlike most other HERVs, the HERV-K (HML-2) proviruses seem to be an exception, since they have been shown to contain open reading frames (ORFs) for *gag*, protease (*prt*), polymerase (*pol*), and envelope (*env*); encode enzymatically functional retroviral proteins; and even produce retrovirus-like particles (17).

The HERV-K superfamily has been classified by alignments of short sequences from a conserved region within the retroviral reverse transcriptase (21), and recent work has suggested the existence of up to 10 of these HML (human endogenous MMTV-like) subgroups (2). The HERV-K family seems to have integrated into the germ line about 30 million years (Myr) ago, prior to the evolutionary split of hominoids and lower Old World monkeys. However, there is also evidence for an ongoing amplification of HERV-K(HML-2) sequences specifically in the hominoid and human lineage, and even human-specific elements have been identified (20, 23). HERV-K(HML-2) proviral sequences exist in two different types, distinguished by a 292-bp deletion at the boundary of the *pol* and *env* genes. Such mutated proviruses are regarded as deficient, since they do not encode a correct Env protein due to the lack of a signal peptide, for instance (27). Since HERV-K genomes harboring this deletion can be detected only in hominoid species, the mutational event appears to have occurred in a hominoid predecessor species after the evolutionary split from lower Old World monkeys. Both types of HERV-K genomes amplified in the hominoid lineage and seem to have contributed equally to the family's copy number in humans (18, 19).

In our previous study on the evolution of HERV-K homologous sequences in Old World primates, we observed a second mutational event that emerged at the same time in evolution, leading to a shortened *gag* gene (19). This deletion of 96 bp shortens *gag* but maintains the ORF and can be found exclusively in hominoid species, whereas the longer *gag* is present in lower Old World primates. Interestingly, only the deleted *gag* sequences seem to have contributed to the amplification and expansion of HERV-K(HML-2) homologues within the hominoid lineage, and the more ancient *gag* variant was apparently not conserved during evolution. Therefore, one hypothesis was that the shortened Gag protein may have acquired an alternate function, perhaps becoming beneficial for the host. The expression of the variant provirus would have resulted, as an indirect

* Corresponding author. Mailing address: Institut für Humangenetik, Bau 60, Universität des Saarlandes, 66421 Homburg/Saar, Germany. Phone: 49-6841-1626038. Fax: 49-6841-1626186. E-mail: hgemee@med-rz.uni-sb.de.

consequence, in retrotransposition, and therefore amplification of proviruses with *gag* ORFs. At the time of our previous report, only one short *Homo sapiens* GenBank entry (accession no. Z58084) showed similarities to the ancient 96-bp *gag*-sequence, which we interpreted as an evolutionary remnant of the older HERV-K(HML-2) genome (19).

Utilizing the dramatically increasing data from the Human Genome Mapping Project (HGMP), we set out to further clarify the evolution of HERV-K(HML-2) sequences. Our analyses led to the identification and characterization of 10 previously unknown ancient HERV-K(HML-2) homologues, with characteristic features of long-time integration that are conserved in the human genome until today.

MATERIALS AND METHODS

BLAST searches and sequence analysis. The GenBank human nonredundant (nr) and high-throughput genomic sequence (htgs) databases were screened by BLAST search (1) with a 96-bp *gag*-sequence from *Cercopithecus aethiops* (accession no. AF018153) characteristic of an ancient HERV-K *gag* (19). To characterize proviral portions in the entries, comparison of the identified sequences with an intact HERV-K element, HERV-K(HML-2.HOM) (20, 28), was done by dot matrix analysis with MacVector software (Genetics Computer Group). Parameters for the dot matrices were a window size of 30 nucleotides (nt) and a minimum similarity of 70%. Proviral fragments identified in the unfinished htgs entries were only subjected to further analyses when they were located in a single contig of the htgs entry. Pairwise sequence comparisons between the newly identified HERV-K(OLD) with one another and also to HERV-K(HML-2.HOM) were performed using BLAST 2 sequences at the National Center for Biotechnology Information (32) and Sequencher (Gene Codes Corporation), which also served in the analysis of potential ORFs. Repetitive elements were identified using the RepeatMasker Web Server (A. F. A. Smit and P. Green, unpublished data [http://ftp.genome.washington.edu/cgi-bin/RepeatMasker]), which was also helpful in the exact localization of HERV-K(OLD) proviral fragments.

Evolutionary ages of proviruses. The approximate integration times of HERV-K(OLD) proviruses were estimated by two different approaches. First, the age of the *Alu* subfamily members (9, 24) inserted into some of the proviruses gave an estimate of the minimum age of the respective proviral element. Second, the evolutionary age of those proviruses with both flanking full-length long terminal repeats (LTRs) was estimated by the sequence comparison of the 5′ and 3′ LTRs. Both LTRs are supposed to be identical in sequence at the time of integration and start to acquire mutations afterward. Indels were excluded in the calculation of the percent divergence between the two LTRs, and the divergence values were then corrected for revertant and superimposed changes (10). The approximate integration time T was obtained by the method given in Lebedev et al. (12), with the formula $T = D/(2 \times 0.13)$, where D is the corrected divergence value and 0.13 is the average mutation rate per Myr for the evolution of LTRs. The factor 2 accounts for the fact that both LTRs acquire mutations independently, so that the sum of mutations in both LTRs contributes to the percentage of divergence.

Sequence alignments and phylogenetic analyses. Multiple sequence alignments of HERV proviral portions were done using the CLUSTALW algorithm (33) at the Institut Pasteur (http://bioweb.pasteur.fr/seqanal/interfaces/clustalw.html). The LTR and *pol* sequences of the proviruses reported here were compared to HERV sequences previously described by others (11, 21, 23). Neighbor-joining, maximum likelihood, and maximum parsimony trees were constructed using PHYLIP (Phylogeny Inference Package, version 3.5c; J. Felsenstein, Department of Genetics, University of Washington, Seattle). Gaps in alignments were excluded from analyses. To obtain support values, bootstrap analysis was performed with 1,000 replicates. The bootstrap values represent the percentage of trees for which the sequences at one end of the branch form a monophyletic group. To establish a HERV-K(OLD) dUTPase consensus sequence from a multiple sequence alignment, the server of the European Molecular Biology Laboratory (http://www.bork.embl-heidelberg.de:8080/Alignment/consensus.html) was used.

PCR primers and conditions. HERV-K(OLD-AL035587) was amplified from total genomic DNA from various primate species using the primers LTRF (5′CTGGCCTATGTGCACATCCAGG3′) and OLD3′FlankR (5′CCAGTCTG GAGGAAACTGACC3′). PCR cycling conditions were as follows: an initial cycle of 5 min at 94°C; followed by 30 cycles of 30 s at 94°C, 30 s at 57°C, and 1 min at 72°C; and a final cycle of 10 min at 72°C. Reaction mixtures contained 0.5 μM concentrations of each primer, 200 μM deoxynucleoside triphosphates, and 2.5 U of *Taq* polymerase in standard PCR buffer (Life Technologies).

RESULTS

Identification of human sequences homologous to the 96-bp *gag*-sequence. The 96-bp sequence characteristic for the HERV-K(HML-2) *gag* genes of lower Old World primates (19) was utilized to perform BLAST searches on the nr and htgs divisions of GenBank, the latter one covering unfinished sequences of the HGMP. As of July 2000, we identified 26 human sequence entries in the nr database and nine entries in the htgs database producing significant E-values. Several of these entries showed a more or less full-length similarity to the query sequence, but some were only similar to the central part of the 96-bp *gag* sequence (Table 1).

The presence of additional proviral portions within the identified sequence entries was examined by dot matrix comparisons with known HERV-K proviruses. We identified novel proviral sequences in 10 of the database entries, and the new HERV elements were named according to the accession number of the corresponding sequence entry (Fig. 1). The locations and orientations of proviral portions within the respective sequence entries are given in Table 2. For all 10 of the identified HERVs, the sequences immediately downstream of the 5′ LTR have been compared to known tRNA sequences from a database (30). All of the identified primer binding sites showed strong homology to the last 18 nt of $tRNA^{Lys}$ (with a mean divergence of 3.7 nt) and less similarity to other tRNA sequences, supporting the classification as HERV-K (data not shown). The dot matrices grouped all new HERV-K proviruses into the HERV-K(HML-2) family, displaying significant similarity to the most intact member of this family HERV-K (HML-2.HOM) (20, 28) (Fig. 1). In contrast, comparisons to members of the HERV-K(HML-4) (29) or HERV-K(HML-6) (22) family produced clearly fewer and shorter similar regions (data not shown). As deduced from these dot matrix analyses, pairwise sequence comparisons and the chromosomal location as annotated in the GenBank file, the following accession numbers represent identical genomic regions and proviral loci, respectively: AL358753 and AL023753, AC010632 and AC012309, AP000346 and AP000345, and AC006078 and AC004127.

Structure of HERV-K(OLD) proviral loci. The structure of the HERV-K(HML-2)-like sequences reported here ranges from almost complete proviruses to the remains of proviruses, with large deletions or rearrangements, and the retroviral ORFs are frequently interrupted by stop codons. Furthermore, *Alu* elements from various *Alu* subfamilies are frequently inserted into the proviral sequences. However, all proviruses display the typical ancient *gag* gene, having the uniform 96-bp sequence. Since all of these features strongly suggest a long-time presence in the human genome, we named this proviral group HERV-K(OLD).

Here we present the 10 proviral loci showing reasonably intact structures in more detail, arranged in the same order as in the BLAST results in Table 1. HERV-K(OLD-AC004979) shows a deletion of the 5′ portion of *pol* as well as the complete *env* gene and the 3′ LTR, respectively. The remaining *gag*, *prt*,

TABLE 1. Search results of the BLASTN query with the 96-bp *gag*-sequence from *Cercopithecus aethiops*[a]

Accession no.	Alignment[b]		
	nt	Sequence	nt
Query	1	caagtccaaaccccaagagaatatcaaatagagaaag-atagagtctctgtcacggcaatgccaatccaaatacagtatccacaataacagtcggta	96
AC004979	8786	.....a.....-.....t...c.....	8691
AC012309	27691	.....a.....a.....g.....-.....t...c....	27597
AL121932	78631	.....a.....ga.....a....-...a.....c..t.....g.....	78716
AC004034	20623	.....a.....a..g.--....-.....a..t.....g.....	20540
AL035587	49498	.....a.....g....a..g.-.....a..c.....t..g..g.....	49583
AL023753	11745	.....g.....g....a..g.-.....a..c.a.....t..g.....	11830
AL031668	26013	.....a.....g.....a..g.--....-.....a....a.....g.....t.....	25930
AP000346	22616	.....a.....g....a.gg.-.c.....a..c.....ct..g.....t...c.....	22705
AP000345	60027	.....a.....g....a.gg.-.c.....a..c.....ct..g.....t...c.....	60116
AL136419	115362	.....g.c.....t...c.....	115323
Z58084	227	.....g.....tg...ga....g.....a..c...nc.....	291
AC008062	47628	.....	47647
AC005177	117272	.....-....	117255
AC004924	31115	..-.....	31098
AC005326	10650	..-.....	10633
AC002394	97019	..-.....	97036
AF019413	90066	..-.....	90049
AC006026	42989	..-.....	43006
AC006925	92637	..-.....	92620
AL109827	509	..-.....	526
AL034548	39526	..-.....	39509
L09706	10251	..-.....	10268
AL035698	196708	.....	196725
AL078583	63078	..-.....	63061
AL031132	111078	.....a.....-..	111099
AP000502	8371	..-.....	8354
AC010632	110491	.....a.....a.....g.....-.....t...c....	110397
AL358783	167647	.....g.....g....a..g.-.....a..c.....t..g.....	167732
AC006078	68622	.....a.....--------.....t.....c.....t...c.....	68709
AC004127	36693	.....a.....---- ----.....t.....c.....t...c.....	36780
AC022412	79909	.....a.....tg.....g....a..g.-.....a..c.....g.c.....	79824
AC024690	138073	.....g.....tg....a..g.-.....a..c....c.....t..g.....	137988
AL355987	48461	.....a.....a.....	48430
AC012068	19924	.....g.c.....	19895
AC068021	59166	.....	59184

[a] Upper part, search results from human nr database; lower part, search results of htgs database (sequences of the htgs databank may now be of finished status). The order of the matches is according to the E-value of the BLAST search.

[b] Numbers indicate the nucleotide (nt) position of the matches within the sequence entry. Dots indicate identities, and dashes indicate insertions compared to the query sequence.

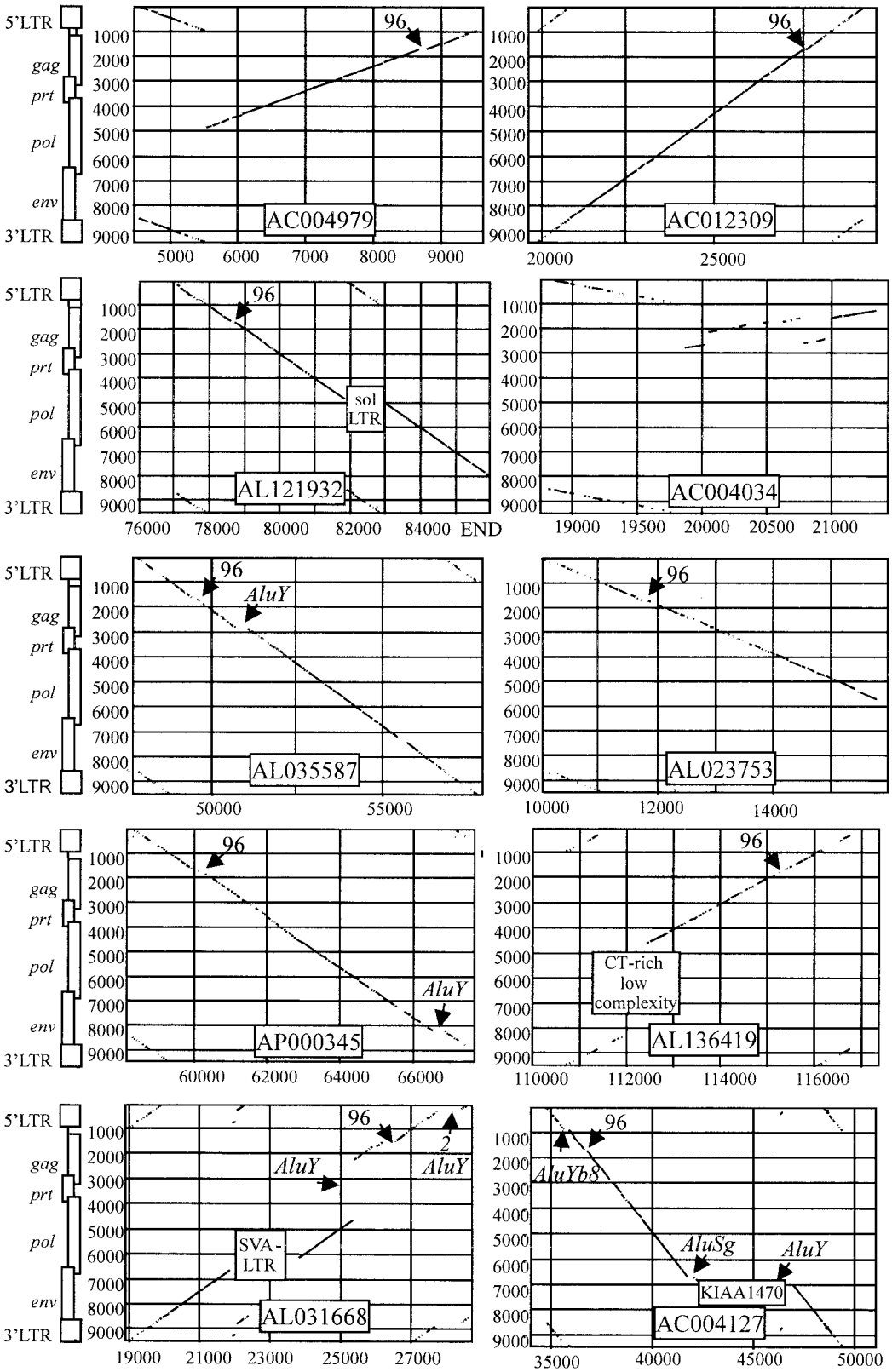

FIG. 1. Dot matrix comparisons between the most intact member of the HERV-K(HML2) family (20, 28) and the identified HERV-K(OLD) proviruses. Lines represent regions with at least 70% similarity in a window of 30 nt between the two proviral sequences. HERV-K(OLD) sequences are shown on the *x* axis, and the numbers give the position in the GenBank entry. The location of features like deletions and insertions are indicated in the matrices.

TABLE 2. Characterization of HERV-K(OLD) proviral elements

Accession no.	Chromosomal location	Position in sequence entry[a]	Identified *Alu* element	Age in Myr of the *Alu* subfamily (SD)[b]	LTR subgroup[c]	% Divergence between 5′ and 3′ LTR[d] (estimated integration date in Myr[e])
AC004979	7q31.3–q32	4543–9556 (−)			−8/−23	
AC012309	19	19871–29390 (−)			−8/−23	7.72 (29.7)
AL121932	6	77077–85952 (+)			−8/−23 (solLTR −8/+23)	
AC004034	16	18813–21376 (+)			+8/+23	
AL035587	6p12.3–21.2	47775–57733 (+)	*AluY* in *gag*	19 (12)	+8/+23	3.61 (13.9)
AL023753	1p36.11–36.31	10026–15846 (+)			+8/+23	
AP000345	22q11.2	58296–67462 (+)	*AluY* in *env*	19 (12)	+8/+23	
AL136419	14	110721–116801 (−)			+8/+23	
AL031668	20q11.21–11.23	19015–28651 (−)	2 × *AluY* in 5′ LTR	19 (12)	+8/+23	5.83 (22.4)
			AluY in *gag*	19 (12)		
AC004127$_{(htgs)}$	11	34788–49392 (+)	*AluYb8* in 5′ LTR	3 (3)	−8/−23	6.57 (25.3)
			AluSg in *pol*/*env*	31 (12)		
			AluY in pseudogene	19 (12)		

[a] + and −, designates orientation in sequence entry
[b] Average age of *Alu* subfamilies in Myr as estimated from divergence values; standard deviations are given in parentheses (9).
[c] See Fig. 2.
[d] Divergence values are corrected for revertant and superimposed changes according to the method of Kimura (10).
[e] The average mutation rate for the evolution of LTRs of 0.13%/Myr was used for the calculation of the integration time by the formula $T = D/(2 \times 0.13)$, with D being the divergence value (12).

and *pol* portions are inverted compared to the 5′ LTR. HERV-K(OLD-AC012309) appears to be the most intact provirus, with a length of 9,519 bp (Fig. 1). A 5′ LTR of 957 bp and a 3′ LTR of 968 bp are present, and the sequence has the highest overall homology to HERV-K(HML-2.HOM). The provirus is integrated into an *AluSx* element. The central region of the *pol* gene in HERV-K(OLD-AL121932) contains a solitary LTR (974 bp) with strong similarities to the HERV-K(OLD) LTRs. A target site duplication AGATCCC for this integration can also be found. This LTR probably results from the integration of an HERV-K element, and a subsequent homologous recombination led to the loss of the proviral sequence, with only a solitary LTR remaining. Approximately 100 bp of the 5′ end of the proviral 5′ LTR appear to be deleted. There is no information on the 3′ end of *env* and the 3′ LTR, since the provirus is located at the end of the GenBank sequence entry. The HERV-K(OLD-AC004034) locus displays the highest degree of mutations, consisting only of a LTR (986 bp) and a *gag* gene. The remaining *gag* portions, showing larger deletions and rearrangements, are located in reverse orientation compared to the 5′ LTR. The HERV-K(OLD-AL035587) provirus has LTRs of 997 and 996 bp, respectively. It appears to be a complete provirus except that an *AluY* element is inserted within the *gag-prt* boundary. HERV-K(OLD-AL023753) is only 5,818 bp long, displaying deletions of the 3′ end of the *pol* gene, the complete *env* gene and the 3′ LTR, respectively. Approximately 2.5 kb downstream of the provirus, the gene for a protein termed "melanoma preferentially expressed antigen" (PRAME) has been annotated in antisense orientation in the sequence entry. Recently, Tristem (34) described another proviral element in the entry AL023753, belonging to the highly defective HERV.HS49C23 family, which is not identical but rather located next to the HERV-K(OLD) provirus described here. Besides some smaller deletions, the HERV-K(OLD-AP000345) provirus has an almost intact proviral structure, with a partially deleted 3′ LTR and an insertion of an *Alu*Y element in the 3′ part of the *env* gene. The proviral sequence HERV-K(OLD-AL136419) has a deletion of 300 bp in the 5′ LTR, compared to the 3′ LTR of 1,024 bp. A short deletion in the *gag* gene explains why there is no full-length similarity to the 96-bp sequence in the initial BLAST search (see Table 1). The 5′ half of the *pol* gene and most of the *env* gene are deleted, and a 480-bp CT-rich low complexity region can be identified here instead. HERV-K(OLD-AL031668) is integrated into an L1 element and itself shows insertions of four additional repetitive elements, including three *AluY* and one SVA element. Two *AluY* elements in tandem orientation are inserted into the 5′ LTR, which is 995 bp without *Alu* elements, compared to the 3′ LTR of 984 bp. Another *AluY* element is located in the *gag* gene. The SVA element, which is a composite retroposon consisting of *Alu* fragments and an HERV-K LTR, is integrated into the *pol-env* boundary. A deletion furthermore removed the region ranging from the 3′ half of *gag* to the 5′ region of *pol*. A provirus found in the htgs section of GenBank, HERV-K(OLD-AC004127), also shows common insertions of other repetitive elements. An *AluYb8* element is inserted into the 5′ LTR, which is 970 bp without the *Alu* element, compared to the 3′ LTR of 960 bp. An *AluSg* element is integrated, similar to the AL031668 provirus, into the *pol-env* boundary. The sequence inserted into the *env* gene (target site duplication CCAGGATG) shows a 94% similarity to the mRNA for a KIAA1470 protein (accession no. AB040903). This presumably processed pseudogene sequence is in the opposite orientation to the provirus and itself contains other repetitive elements, such as AluY, SVA-LTR, and simple repeats. Two other htgs entries, AC024690 and AC022412, also contain relatively intact HERV-K sequences, as deduced from dot matrices, but are due to the unfinished status of the sequence assembly not considered here.

Most genes in the HERV-K(OLD) proviruses have accumulated multiple stop codons, but sometimes longer ORFs could be detected, e.g., a 623-amino-acid (aa) ORF for *gag* in

```
                612                                                                                                  668
5'HML-2.HOM     TACTGCTTT--------GTAAAGCACTGAGATGTTTATGTG----------------------TATGCATATCTAAAAGCACAGCAC
3'HML-2.HOM     TACTGCTTT--------GTAAAGCACTGAGATGTTTATGTG----------------------TATGCATATCTAAAAGCACAGCAC
5'AC012309      TACTGCTAT--------TTAATGCACCGAGATGTTTGTATA----------------------CATGTACATC--AAGGCACAGCAC
3'AC012309      TACTGCTCT--------TTAATGCACCGAGATGTTTGTACA----------------------CGTGCACATC--AAGGCACAGCAC
5'AC004127      TACTGCTCT--------TTAATG----------TTTGTATA----------------------CATGCACATC--AAGGCATAGCAC
3'AC004127      TACTGCTCT--------TTAATGCACCAAGATGTTTGTATA----------------------CGTGCACATC--AAGGCACAGCAC
5'AL121932      TATTGCTCT--------TTAATGCACTGAGGTGTTTGTGTA----------------------CCTGCATATC--AAAGCACAGCAC
solAL121932     TACTGTTCC--------TTAATGCACCAAGATGTTTGTGTAAAGTCAAACATAAATCTGGCCTATGTGCACATC--AAGGCACAGCAC
3'AC004979      TACTGCTCT--------TTAAGGCATTGAGGTGTTTACATA----------------------TGTGCACATC-AAAAGCACAGCAC
5'AL031668      TACTGCTCTGTTACTCTTTACTGCACTGAGGTGTTTATGTAAGCTTAAACATAAATCTAGCGATTGTGCACATCC--AGGCACAGCAC
3'AL031668      TACTGCTCTGTTACTCTTTACTGCACTGAGTTGTTTATGTAAGCTTAAACATAAATCTAGCGATTGTGCACATCC--AGGCACAGCAC
5'AL035587      TACTGCTCTGTTATTCTTTACTACACTGAGATGTTTGGGTGGAGAGAAGCATGAGTCTGGCCTATGTGCACATCC--AGGCATAGTAC
3'AL035587      TACTGCTGTGTTATTCTTTACTACACTGAGATGTTTGGGTGGAGAGAAGCATATATCTGGCCTATGTGCACATCC--AGGCATAGTAC
5'AP000345      TGCTGCTCTGTTACTCTTTGCTACACTGAGATGTTTGGGTGGAGAGAAGCATAAATCTGGCCTATGTGCACATCT--GGGCACAGAAC
5'AL136419      TGCTGCCTTGTTATTCTTTACTCCACTGAGATGTTTGGGTGGAGAGAAACATAAATCTGGCTTACATGCACGTCC--AGTCATAGTAC
3'AL136419      TGCTGCCTCGTTATTCTTTACTCCACCGAGATGTTTGGGTGGAGAGAAACATAAATCTGGCTTACGTGCATGTCC--AGTCATAGTAC
5'AL023753      TGCTGCTTTGTTACTCTTTACTCCACTGAGATGTTTGGGTGGAGAAAAGCATAAATCTGGCCTATGTGCACATCC--AGGCATAGTAC
5'AC004034      TACTGCTCTGTTACTCTTTGCTACACTGAGATGTTTGTGTGAAGTGAAACACAAATCTGGCCTACATGCACATCC--AGACACAGTAC
```

FIG. 2. Multiple sequence alignment of HERV-K(OLD) LTR sequences. Only the LTR regions harboring the diagnostic differences between the two subgroups are shown. Positions 612 to 668 of the human-specific HERV-K(HML-2.HOM) LTRs (20, 28) are shown for comparison. 5′ and 3′ indicates the respective LTRs in a given provirus. "solAL121932" indicates the solitary LTR present in HERV-K(OLD-AL121932).

HERV-K(OLD-AP000345). However, a full-length ORF for the more ancient *gag* genes is estimated ca. 730-aa. The *prt* ORF in HERV-K(OLD-AL136419) is intact, extending into *pol* and encoding 321 aa. Nevertheless, a potential for protein expression from these ORFs is questionable.

Evolutionary age of HERV-K(OLD) proviruses. The age of several *Alu* subfamily members, inserted into some of the HERV-K(OLD) loci, can be used to estimate the minimal age of the retroviral elements (Table 2). Since *Alu* families have relatively high ranges of evolutionary ages (9, 24), these data give only an approximate estimate of integration times. For HERV-K(OLD-AC004127), the minimum time of its presence in the genome can also be calculated from the sequence divergence of the KIAA1470 pseudogene compared to the mRNA (26). Based on a mutation rate of 1.5×10^{-9} substitutions/site/year for nonfunctional sequences (15) and the 6% divergence to the KIAA1470 mRNA sequence, the pseudogene, was formed roughly 20 Myr ago, when the provirus was already present.

At the time of the proviral integration, both LTRs of a retrovirus are identical in sequence but then acquire mutations independently over time, thus providing an evolutionary clock for the age of the provirus (5). For those HERV-K(OLD) proviral elements with apparently complete 5′ and 3′ LTRs, the corrected LTR sequence divergence was used to calculate approximate integration times (Table 2). Taken together, our data hint at HERV-K(OLD) proviral integrations around the evolutionary split of lower Old World primates and the hominoid lineage, which occurred ca. 28 Myr ago (4, 31).

HERV-K(OLD) LTR sequences. A sequence alignment of HERV-K(OLD) and the human-specific "modern" HERV-K(HML-2.HOM) LTR sequences led to the distinction of two HERV-K(OLD) subgroups. The LTRs of one subgroup proved to be longer in sequence, whereas the other consistently displayed deletions of 8 and 23 bp, and were therefore similar to HERV-K(HML-2.HOM) (Fig. 2). Additionally, LTRs of both subgroups showed a deletion of 2 bp compared to the HERV-K(HML-2.HOM) LTR. Interestingly, the solitary LTR inserted into the *pol* gene of HERV-K(OLD-AL121932) is an intermediate between the two subgroups, since it shows only the first deletion of 8 bp.

HERV-K(HML-2) LTR sequences have been previously reported to be grouped into evolutionary older and younger clusters (11, 12). We performed a neighbor-joining analysis of the HERV-K(OLD) LTR sequences and included some representatives of these older and younger HERV-K LTRs. As can be seen in Fig. 3 the HERV-K(OLD) LTRs cluster with LTRs of the evolutionary older group (LTRI) and are separated from the younger LTRII sequences with a 100% bootstrap support. Furthermore, HERV-K(OLD) LTRs without the described deletions of 8 and 23 bp cluster separately from those containing these deletions, which appear phylogenetically more related to the younger LTRII group. The solitary LTR in HERV-K(OLD-AL121932) has an intermediate position in the tree.

Phylogenetic relationship of HERV-K(OLD) polymerase sequences. Previously, the HERV-K sequences in the human genome have been divided into at least six subfamilies, HML-1 through HML-6, based on sequence comparison of conserved reverse transcriptase regions (2, 21). To examine the phylogenetic relationship between HERV-K(OLD) and other HERV-K families, we compared the HERV-K(OLD) *pol* regions present in eight proviruses with these previously described HML *pol* sequences. Additionally, the *pol* sequences of the proviruses HERV-K(HML-2.HOM) (20, 28) and HERV-K-T47D (29) have also been included. The phylogenetic analyses resulted in similar tree topologies for the neighbor-joining, maximum likelihood, or maximum parsimony methods. Remarkably, two HERV-K(OLD) *pol* sequences proved to be identical to the described HML2-4 and HML2-5 *pol* sequences (Fig. 4). The neighbor-joining tree in Fig. 4 shows that all HERV-K(OLD) *pol* sequences cluster with a bootstrap support of 100% with HERV-K(HML-2) sequences. However, the various HML-2 and HERV-K(OLD) sequences did not cluster

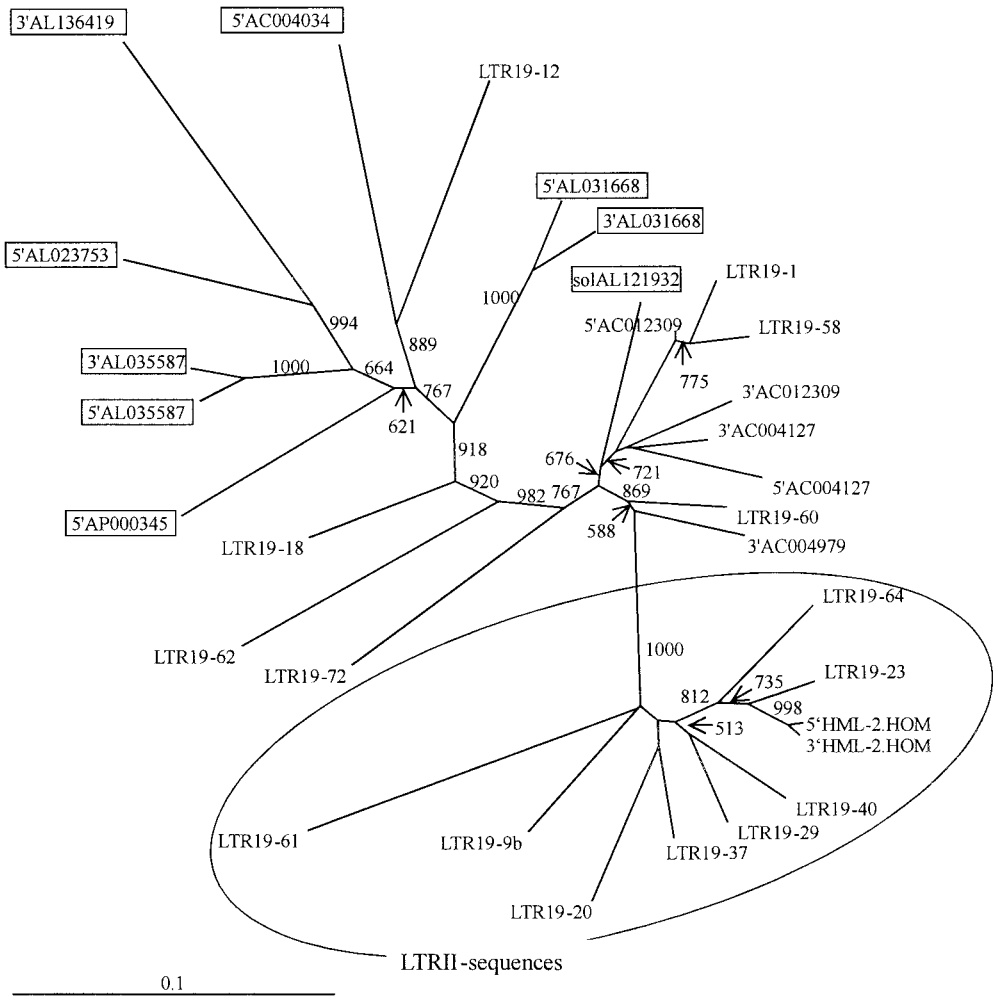

FIG. 3. Unrooted neighbor-joining tree of HERV-K LTR sequences. HERV-K(OLD) LTRs were compared to previously reported LTR sequences from evolutionary older (LTRI) and younger (LTRII) clusters derived from chromosome 19 (11). HERV-K(OLD) LTRs without the diagnostic deletions of 8 and 23 bp are boxed, and LTRII sequences are circled. All HERV-K(OLD) LTRs cluster with the older LTRI sequences (not included in the circle), independent from the presence or absence of the 8- and 23-bp deletions (see text for further details). The phylogenetic distances were calculated using the Kimura two-parameter model with a transition/transversion ratio of 2. Support values (1,000 bootstrap replicates) are indicated on the corresponding branches. The scale bar represents a 10% evolutionary distance.

as a single group but formed two subgroups with bootstrap supports of 72% for both clades. Interestingly, the one cluster contains all HERV-K(OLD) sequences for which longer LTRs have been determined, while the other one contains the proviruses having the shorter LTR variant, thus confirming the distinction of the two subgroups.

Analysis of the HERV-K(OLD) dUTPase region. HERV-K elements encode in the *prt* reading frame a rudimentary dUTPase. dUTPase catalyzes the hydrolysis of dUTP to dUMP and PP_i in some retroviruses (7), and it has been shown that dUTPase might represent an example of horizontal gene transfer between eukaryotic and archaeal hosts and viral pathogens (3). However, the dUTPase protein sequence of the human-specific HERV-K(HML-2.HOM) (20) shows substitutions at highly conserved residues in four of five motifs typically well conserved in dUTPase enzymes and therefore likely contributes to an nonfunctional enzyme. The HERV-K(OLD) *prt* sequences are relatively intact, in that they contain only a few stop and frameshift mutations; a *prt* ORF can even be found in HERV-K(OLD-AL136419). Therefore, we compared an HERV-K(OLD) *prt* consensus sequence, derived from an alignment of eight HERV-K(OLD) sequences, to the corresponding dUTPase motifs of reported active enzymes (Fig. 5). The HERV-K(OLD) *prt* consensus sequence displayed an ORF, and the translated amino acid sequence resulted in motifs resembling those of a reported enzymatically active HERV-K(HML-2) dUTPase construct (7). Most notably the residues Gly (motif 3) and Phe (motif 5), which are important for substrate recognition (25), and the residue Gly (motif 4) seem to be wild type-like. Thus, HERV-K(OLD) proviruses may have encoded an active dUTPase.

DISCUSSION

Previously, we reported that HERV-K(HML-2) homologous sequences in lower Old World primates exhibit a 96-bp-

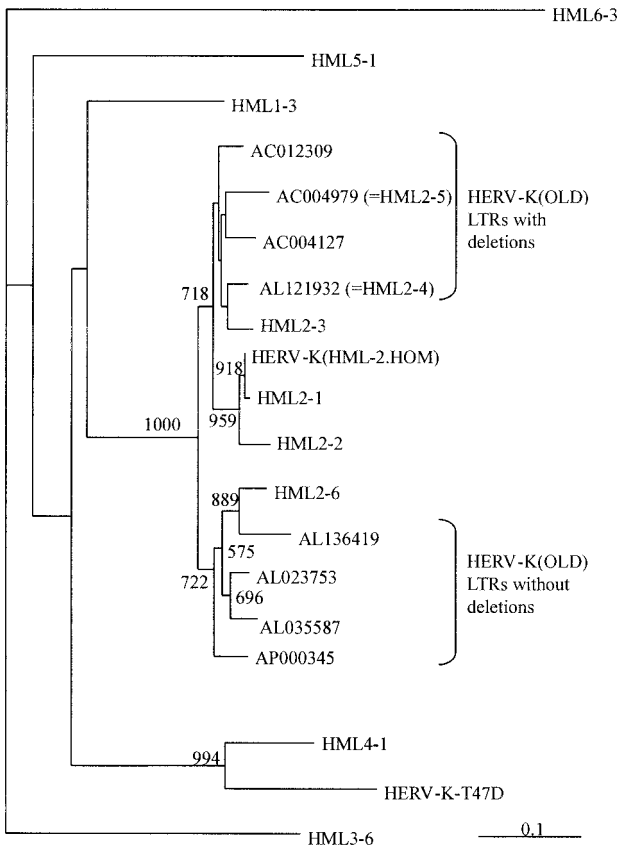

FIG. 4. Neighbor-joining tree of HERV-K(OLD) polymerase sequences and previously published HML polymerase sequences (21). The tree is rooted with a HML-6 sequence being the most distant of the six HML groups (21), but the HML6-3 and HML3-6 relation could not be resolved in this analysis. The distances were calculated using the Kimura two-parameter model with a transition/transversion ratio of 2. Support values (1,000 bootstrap replicates) are indicated on the corresponding branches. The scale bar represents a 10% evolutionary distance.

longer *gag* gene compared to human sequences. The shorter *gag* gene obviously arose after the evolutionary split of hominoids from lower Old World primates and is amplified in copy number. The more ancient HERV-K(HML-2) *gag* sequences could not be clearly identified in the human sequence data available at that time (19). With the increased sequence data generated by the HGMP, we were now able to identify the remnants of such ancient HERV-K(HML-2) homologous proviruses. The 10 detected retroviral elements show multiple signs of long-time presence in the genome. The ORFs are usually interrupted by multiple stop codons, deletions and inversions result in the loss or rearrangement of large proviral portions, and some proviruses also contain *Alu* elements from various families. We therefore named these retroviral sequences HERV-K(OLD).

Using the age of inserted *Alu* subfamily members and the calculated divergence values between 5′ and 3′ LTRs of HERV-K(OLD) proviruses, it was possible to roughly estimate the age of some retroviral elements. For instance, the LTR divergence in HERV-K(OLD-AC012309) indicates a presence in the genome for ca. 30 Myr. LTR divergence and the presence of an *AluSg* element in HERV-K(OLD-AC004127) date this provirus from ca. 25 and 31 Myr ago, respectively. It must be taken into account that the calculation of integration times from LTR divergences assumes rate constancy, and thus a molecular clock in the primate lineage but a slowdown in the mutation rate in the hominoid lineage is conceivable (14). Attempts were made to experimentally address the age of single HERV-K(OLD) proviruses, but PCR studies proved to be technically difficult, due to multiple mutations in proviral and flanking regions, preventing effective primer design. Nevertheless, in one case it was possible to trace the AL035587 provirus back to lower Old World Monkeys (Fig. 6), although the presence of an *AluY* element dates this provirus to only 19 Myr ago. This supports the notion that the calculated integration times given above yield only minimum ages, and thus the examined endogenous retroviruses may be considerably older. Another hint at the evolution of these proviruses is the presence of a 292-bp sequence at the *pol-env* boundary in six of the identified proviruses. (The remaining four proviruses lack this sequence due to larger deletions within the *pol* or *env* genes; see Fig. 1.) HERV-K proviral elements retaining this 292-bp sequence are thought to be the more ancient variants (16, 27). We therefore suggest that the HERV-K(OLD) elements are remnants from a time ca. 28 Myr ago when the genomes of lower Old World primates and hominoids were not yet evolutionarily separated (4, 31).

	dUTPase-motif 1	2	3	4	5
Human	AGYDL	RSGLAAK	GVIDEDYRG	GDRIAQLI	RGSGGFGSTG
Herpes simplex	AGYDI	RSSLNAR	GLIDSGYRG	GAKVAQLV	RGTRGFGSTG
MMTV	AGLDL	RSSNYKK	GVIDSDFQG	GERIAQLI	RGSEGFGSTS
HERV-K(HML-2.HOM)	A**A**VDL	RSSLNLK	**S**VVDSDYKG	**R**DRIAQLL	**KR**IGG**L**GSTD
HERV-K Cons	A**A**VDL	RSSLNLK	GVVDSDYKG	GDRIAQLL	**KR**IGGFGSTD
AL136419	A**A**VDL	RSSLNLK	GVIDSDYKG	GDRIAQLL	**ER**TGGFGSTN
HERV-K(OLD) Cons	A**A**VDL	RSSLNLK	GVIDSDYKG V	GDRIAQLL	**ER**TGGFGSTD N

FIG. 5. Amino acid alignment of conserved dUTPase motifs. Residues in HERV-K proviral sequences differing from highly conserved residues are indicated in boldface. "HERV-K Cons" corresponds to an enzymatically active HERV-K(HML-2) dUTPase construct (7). "HERV-K(OLD) Cons" is the dUTPase consensus sequence derived from eight proviral sequences. Ambiguous amino acids are indicated below the sequence. The amino acid sequence of HERV-K(OLD-AL136419), displaying a *prt* ORF, is included for comparison.

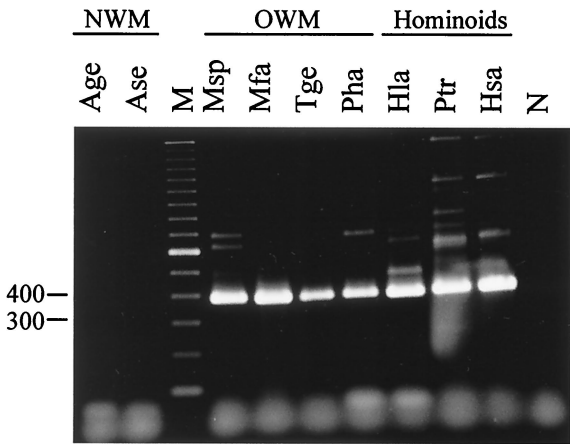

FIG. 6. Presence of HERV-K(OLD-AL035587) in hominoids and lower Old World monkeys. PCR primers were specific for 3′-flanking sequences and the HERV-K(OLD) LTR. NWM (New World monkeys): Age, *Ateles geoffroyi*; Ase, *Alouatta seniculus*. OWM (Old World monkeys): Msp, *Mandrillus sphinx*; Mfa, *Macaca fascicularis*; Tge, *Thercopithecus gelada*; Pha, *Papio hamadryas.* Hominoids: Hla, *Hylobates lar*; Ptr, *Pan troglodytes*; Hsa, *Homo sapiens*. M, marker (size of marker bands is given in base pairs); N, PCR control without DNA.

Our study reveals further differences in these ancient HERV-K(HML-2) homologous sequences within the LTR R-region. While one subgroup of HERV-K(OLD) displays LTRs similar to modern HERV-K(HML-2), the other subgroup harbors longer LTRs, with insertions of 8 and 23 bp, respectively, within the LTR R-region. Interestingly, the solitary LTR in HERV-K(OLD-AL121932) is an intermediate between both LTR variants, lacking the 8-bp sequence and containing the 23-bp sequence. A similarly structured solitary LTR can also be found in the HERV-S71 provirus (13, 35). It seems possible that such partially deleted LTRs are the evolutionary intermediates between the obviously more ancient longer LTR variants and the shorter LTR variants. However, the biological consequence of these sequence variations remains to be elucidated.

Two recent reports investigated the phylogenetic relationships of HERV-K(HML-2) LTR sequences (11, 23), and it was now possible to link these sequences to HERV-K(OLD) proviruses. First, Lavrentieva et al. (11) defined LTR groups I and II; the LTRI group was considered evolutionarily older. Our phylogenetic analysis indicates that all HERV-K(OLD) LTRs cluster with the older LTRI sequences, independent from the presence or absence of the 8- and 23-bp deletions (Fig. 3). Consistently, some of the subgroups defined for LTRI (I-E, D, S, P, and K) do not display the 8- and 23-bp deletions; the LTRI-Y group shows only the first deletion of 8 bp; and all other LTRI and -II groups have both deletions. In concordance with our data, LTRI-Y has been estimated to be 41 Myr, and the LTRI groups without the deletions have been estimated to be between 53 and 25 Myr (12). Second, Medstrand and Mager defined nine evolutionary clusters of HERV-K (HML-2) LTRs (23). Here, only the oldest cluster contains the longer HERV-K(OLD)-like LTRs. LTRs of the younger clusters show both deletions. The integration time of LTRs in the oldest cluster has been estimated to be 45 to 30 Myr (23).

The various HERV-K families in the genome have been defined by phylogenetic analysis of a conserved *pol* region (21). Our phylogenetic analysis (Fig. 4) groups HERV-K(OLD) unambiguously in the HML2-family and also confirms the existence of two subgroups differing in LTR length. Interestingly, the recently described sequences HML2-4 and HML2-5 (21) proved to be identical to the *pol* regions in HERV-K(OLD-AL121932) and HERV-K(OLD-AC004979), respectively. It is possible that these *pol* fragments are derived from the HERV-K(OLD) proviruses. Both HML-2.4 and HML-2.5 belong to HERV-K(OLD) proviruses with LTRs carrying the characteristic deletions. It is conceivable that other HML-2 sequences also belong to not-yet-identified proviruses with or without these mutations, for example, HML-2.3, -2.2, and -2.1 to such proviruses with shorter LTRs and HML-2.6 to a more ancient provirus with longer LTRs.

The dUTPase domain within the *prt* ORF is inactive in HERV-K(HML-2) due to mutations in several, usually conserved, amino acid motifs. Alignment of HERV-K(OLD) *prt* sequences led to a consensus sequence with a dUTPase ORF and characteristic amino acid motifs (Fig. 5) resembling those of an enzymatically active HERV-K dUTPase construct, which was interpreted as the ancestral wild-type variant (7). Furthermore, the provirus AL136419 still shows a *prt* ORF. We therefore suggest that HERV-K(OLD) once encoded an active dUTPase enzyme.

In this study we present the apparent predecessor sequences of the HERV-K(HML-2) family, as they were probably present in the genomes before the divergence of the lower Old World primate and hominoid lineages. Based on the data presented here and on previous results (19), these HERV-K(HML-2) predecessors had a longer *gag* gene, a *pol-env* boundary containing the 292-bp sequence, and either short or long LTR sequences, with the long LTR variant conceivably being more ancient. All of these sequence features were changed in the course of the evolution of the human lineage when mutants were amplified and eventually ended up in the human genome. A similar history has been observed for the HERV-H family (6). As with the HERV-K(HML-2) sequences in the human genome, it is possible that in each primate species a unique subset of endogenous retroviruses amplified in copy number after the species separated from its predecessors. Which factors triggered selective amplification of certain endogenous retrovirus variants remains to be seen.

ACKNOWLEDGMENTS

This work is supported by the Deutsche Forschungsgemeinschaft (Me917/16–1).

We thank Brenda Glass for critical editing of the manuscript.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Andersson, M., M. Lindeskog, P. Medstrand, B. Westley, F. May, and J. Blomberg.** 1999. Diversity of human endogenous retrovirus classII-like sequences. J. Gen. Virol. **80:**255–260.
3. **Baldo, A. M., and M. A. McClure.** 1999. Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. J. Virol. **73:**7710–7721.
4. **Britten, R. J.** 1994. Evidence that most human *Alu* sequences were inserted in a process that ceased about 30 million years ago. Proc. Natl. Acad. Sci. USA **91:**6148–6150.
5. **Dangel, A. W., B. J. Baker, A. R. Mendoza, and C. Y. Yu.** 1995. Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular

clock of evolution. Immunogenetics **42:**41–52.
6. **Goodchild, N. L., D. A. Wilkinson, and D. L. Mager.** 1993. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. Virology **196:**778–788.
7. **Harris, J. M., R. H. Haynes, and E. M. McIntosh.** 1997. A consensus sequence for a functional human endogenous retrovirus K (HERV-K) dUTPase. Biochem. Cell. Biol. **75:**143–151.
8. **International Human Genome Sequencing Consortium.** 2001. Initial sequencing and analysis of the human genome. Nature **409:**860–921.
9. **Kapitonov, V., and J. Jurka.** 1996. The age of Alu subfamilies. J. Mol. Evol. **42:**59–65.
10. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
11. **Lavrentieva, I., P. Khil, T. Vinogradova, A. Akhmedov, A. Lapuk, O. Shakhova, Y. Lebedev, G. Monastyrskaya, and E. D. Sverdlov.** 1998. Subfamilies and nearest-neighbour dendrogram for the LTRs of human endogenous retrovirus HERV-K mapped on chromosome 19: physical neighbourhood does not correlate with identity level. Hum. Genet. **102:**107–116.
12. **Lebedev, Y. B., O. S. Belonovitch, N. V. Zybrova, P. P. Khil, S. G. Kurdyukov, T. V. Vinogradova, G. Hunsmann, and E. D. Sverdlov.** 2000. Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. Gene **247:**265–277.
13. **Leib-Mosch, C., M. Haltmeier, T. Werner, E. M. Geigl, R. Brack-Werner, U. Francke, V. Erfle, and R. Hehlmann.** 1993. Genomic distribution and transcription of solitary HERV-K LTRs. Genomics **18:**261–269.
14. **Li, W.-H., and M. Tanimura.** 1987. The molecular clock runs more slowly in man than in apes and monkeys. Nature **326:**93–96.
15. **Li, W.-H.** 1997. Molecular evolution. Sinauer, Sunderland, Mass.
16. **Löwer, R., K. Boller, B. Hasenmaier, C. Korbmacher, N. Mueller-Lantzsch, J. Löwer, and R. Kurth.** 1993. Identification of human endogenous retroviruses with complex mRNA expression and particle formation. Proc. Natl. Acad. Sci. USA **90:**4480–4484.
17. **Löwer, R., J. Löwer, and R. Kurth.** 1996. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc. Natl. Acad. Sci. USA **93:**5177–5184.
18. **Mayer, J., E. Meese, and N. Mueller-Lantzsch.** 1997. Chromosomal assignment of human endogenous retrovirus K (HERV-K) env open reading frames. Cytogenet. Cell. Genet. **79:**157–161.
19. **Mayer, J., E. Meese, and N. Mueller-Lantzsch.** 1998. Human endogenous retrovirus K homologous sequences and their coding capacity in old world primates. J. Virol. **72:**1870–1875.
20. **Mayer, J., M. Sauter, A. Rácz, D. Scherer, N. Mueller-Lantzsch, and E. Meese.** 1999. An almost-intact human endogenous retrovirus K on human chromosome 7. Nat. Genet. **21:**257–258.
21. **Medstrand, P., and J. Blomberg.** 1993. Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. J. Virol. **67:**6778–6787.
22. **Medstrand, P., D. L. Mager, H. Yin, U. Dietrich, and J. Blomberg.** 1997. Structure and genomic organization of a novel human endogenous retrovirus family: HERV-K (HML-6). J. Gen. Virol. **78:**1731–1744.
23. **Medstrand, P., and D. L. Mager.** 1998. Human-specific integrations of the HERV-K endogenous retrovirus family. J. Virol. **72:**9782–9787.
24. **Mighell, A. J., A. F. Markham, and P. A. Robinson.** 1997. *Alu* sequences. FEBS Lett. **417:**1–5.
25. **Mol, C. D., J. M. Harris, E. M. McIntosh, and J. A. Tainer.** 1996. Human dUTP pyrophosphatase: uracil recognition by a beta hairpin and active sites formed by three separate subunits. Structure **4:**1077–1092.
26. **Nagase, T., R. Kikuno, K. Ishikawa, M. Hirosawa, and O. Ohara.** 2000. Prediction of the coding sequences of unidentified human genes. XVII. The complete sequence of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. **7:**143–150.
27. **Ono, M., T. Yasunaga, T. Miyata, and H. Ushikubo.** 1986. Nucleotide sequence of human endogenous retrovirus genome related to mouse mammary tumor virus genome. J. Virol. **60:**689–698.
28. **Reus, K., J. Mayer, M. Sauter, D. Scherer, N. Müller-Lantzsch, and E. Meese.** 2001. Genomic organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (ERVK6) on chromosome 7. Genomics **72:**314–320.
29. **Seifarth, W., C. Baust, A. Murr, H., Skladny, F. Krieg-Schneider, J. Blusch, T. Werner, R. Hehlmann, and C. Leib-Mösch.** 1998. Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. J. Virol. **72:**8384–8391.
30. **Sprinzl, M., C. Horn, M. Brown, A. Ioudovitch, and S. Steinberg.** 1998. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **26:**148–153.
31. **Takahata, N., and Y. Satta.** 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc. Natl. Acad. Sci. USA **94:**4811–4815.
32. **Tatusova, T. A., and T. L. Madden.** 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. **174:** 247–250.
33. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
34. **Tristem, M.** 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. J. Virol. **74:**3715–3730.
35. **Werner, T., R. Brack-Werner, C. Leib-Mosch, H. Backhaus, V. Erfle, and R. Hehlmann.** 1990. S71 is a phylogenetically distinct human endogenous retroviral element with structural and sequence homology to simian sarcoma virus (SSV). Virology **174:**225–238.
36. **Wilkinson, D. A., D. L. Mager, and J. A. Leong.** 1994. Endogenous human retroviruses, p. 465–535. *In* J. A. Levy (ed.), The Retroviridae, vol. 3. Plenum Press, Inc., New York, N.Y.